

PhUSE 2014

Paper DH09

Challenges of different MedDRA versions for integrated data and databases

Peter Bonata, Bayer Pharma AG, Wuppertal, Germany

ABSTRACT

For the creation of an initial integrated adverse event analyses dataset it is a common situation that the underlying studies are coded in different MedDRA® versions. Therefore a recoding and update to a single version is necessary for a harmonized dataset that assures reliable results and the traceability across versions.

So far “neither the current version of the ADaMIG nor this document fully covers the integration of multiple studies” (CDISC ADaM Data Structure for Adverse Event Analysis Version 1.0^a).

This paper shows a workflow how in cooperation with a standard team these tasks can be fulfilled as well as how former dictionary versions can be handled. In contrast to the CDISC proposal to keep the original coded terms on record level in the data set, it is suggested here to create an analyses dataset history domain containing original and later versions. The advantages of this approach are traceability of different dictionary versions, reproducibility of evaluations and reusability of existing programs when (re)analyzing adverse events.

INTRODUCTION

One of the main tasks for safety evaluation based on an integrated database is the analysis of Adverse Events (AE) as “any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product [.]” (ICH E2A^b). Therefore the Medical Dictionary for Regulatory Activities (MedDRA ^c) is global standard. This coding dictionary is maintained and updated twice a year.

A consequence is, that a process has to be implemented, that ensures during clinical development, for Integrated Summary of Safety (ISS) for submissions and during life cycle management

- a unique thesaurus version across different studies in an integrated database
- allows the traceability of changes,
- the reproducibility of results using former versions is given and
- that datasets are still analyses ready.

When reading the available documentation (ADaMIG ^d) and (draft) guidelines there is no clear procedure outlined how to address this topic for the integration of multiple clinical studies. There are recommendations ^(a) and preparations for a general structure for incidence data as Occurrence Data structure (ODS) in the Analysis Data Model Structure for Occurrence Data^e document. But there are only suggestions outlined that also can be conflicting especially when regarding the data in a life cycle management and not for submission support only.

It is generally stated that the “same single version of MedDRA [should be used] to avoid version specific differences and to correctly identify number of adverse events. The reason for this request is that reviewers often want to analyze adverse events across trials, including the use of Standardized MedDRA Queries. If different

PhUSE 2014

dictionary versions are used for data included in the same analysis, there is the potential for confusion or incorrect results. This is the only way to ensure a truly consistent and coherent comparison of clinical and scientific concepts across multiple studies” (Study Data Technical Conformance Guide¹).

It is also recommended but not required that all levels of terms for the primary path in the MedDRA hierarchy are included (^e).

So it is quite obvious that one single version for the analysis of adverse events in more than one clinical study is expected.

But what are the recommendations for the handling of dictionary coded AE variables in origin and later versions?

The ADAE structure for the standard AE safety dataset is clearly outlined: Keeping multiple sets of mapping variables is not common, but it might be usable when different versions are used for interim and final analyses or when studies are pooled for an integrated analysis. A proposal for providing traceability is to introduce a “counter” in the variable name. This integer “y” from 1 to 9 refers to a previous version. In the metadata the description is enriched by the dictionary name and version.

Table 4.2.10.1 Original or Prior MedDRA Coding Variables

Variable Name	Variable Label	Type	Code List / Controlled Terms	Core	CDISC Notes
DECDORGy	PT in Original Dictionary y	Char	MedDRAy*	Perm	Original preferred term coding of XX.--TERM using MedDRA or other dictionary version X.X.
BDSYORGy	SOC in Original Dictionary y	Char	MedDRAy*	Perm	Original body system coding of XX.--TERM using MedDRA or other dictionary version X.X.
HLGTORGy	HLGT in Original Dictionary y	Char	MedDRAy*	Perm	Original HLGT coding of XX.--TERM using MedDRA or other dictionary version X.X.
HLTORGy	HLT in Original Dictionary y	Char	MedDRAy*	Perm	Original HLT coding of XX.--TERM using MedDRA or other dictionary version X.X.
LLTORGy	LLT in Original Dictionary y	Char	MedDRAy*	Perm	Original LLT coding of XX.--TERM using MedDRA or other dictionary version X.X.
LLTNORGy	LLT Code in Original Dictionary y	Char	MedDRAy*	Perm	Original LLT code of XX.--TERM using MedDRA or other dictionary version X.X.

* For each version of an external dictionary, a different reference name must be used. The individual reference names will point to a dedicated section in the data definition file where all external dictionaries used in the analysis are listed, including dictionary name and version.

This counter is helpful when the continuous updating of MedDRA does not exceed 4 ½ years or nine MedDRA versions. If you have incidence data initially coded 5 years ago a single-digit enumerator is no longer useful. In addition the detachment of an integer to the numeric thesaurus version is not intuitive and might also lead to confusion and complicates the supportive traceability.

It is also stated that a sponsor can create additional analysis datasets for AE analysis, even when using a different structure (^a).

But how does it look when adding more than the primary path to the reported terms?

In this case two counters would be needed to point on the hand side to the secondary (or higher) path and to the different thesauri versions.

But how does it look when Standardized MedDRA Query (SMQ) - as groupings of MedDRA terms at the PT level - or Customized MedDRA Queries (CMQ) variables should be maintained in different versions?

Here is the recommendation to use a number starting from 01 for each query of interest (^e).

PhUSE 2014

Table 4.2.9.1 Standardized MedDRA Query Variables

Variable Name	Variable Label	Type	Code List / Controlled Terms	Core	CDISC Notes
SMQzzNAM	SMQ zz Name	Char		Cond	The standardized MedDRA queries name. Would be blank for terms that are not in the SMQ. Therefore this variable could be blank for all records if no terms within the study were included in the SMQ. Conditional on whether SMQ analysis is done.
SMQzzCD	SMQ zz Code	Num		Perm	The standardized MedDRA queries number code.
SMQzzSC	SMQ zz Scope	Char	BROAD, NARROW	Cond	The search strategy for SMQs can be narrow or broad. The preferred terms that are narrow in scope have high specificity for identifying events of interest while the broad terms have high sensitivity. By definition, all narrow terms are also considered within the broad scope. Therefore, to summarize all broad terms, terms with either narrow OR broad would be considered. Will be null for terms that do not meet the criteria. Conditional on whether SMQ analysis is done.
SMQzzSCN	SMQ zz Scope (N)	Num	1, 2	Perm	Will be null for terms that do not meet the criteria.
CQzzNAM	Customized Query zz Name	Char		Cond	The customized query (CQ) name or name of the AE of special interest category based on a grouping of terms. Would be blank for terms that are not in the CQ. Conditional on whether CQ analysis is done. Examples: "DERMATOLOGICAL EVENTS" "CARDIAC EVENTS", "IARS (INFUSION ASSOCIATED REACTIONS)"

Again the question: How to implement here a reference to the current and prior MedDRA versions? When you take in account that a SMQ or more likely a CMQ can change its specific version within a single MedDRA version then the numbering becomes challenging.

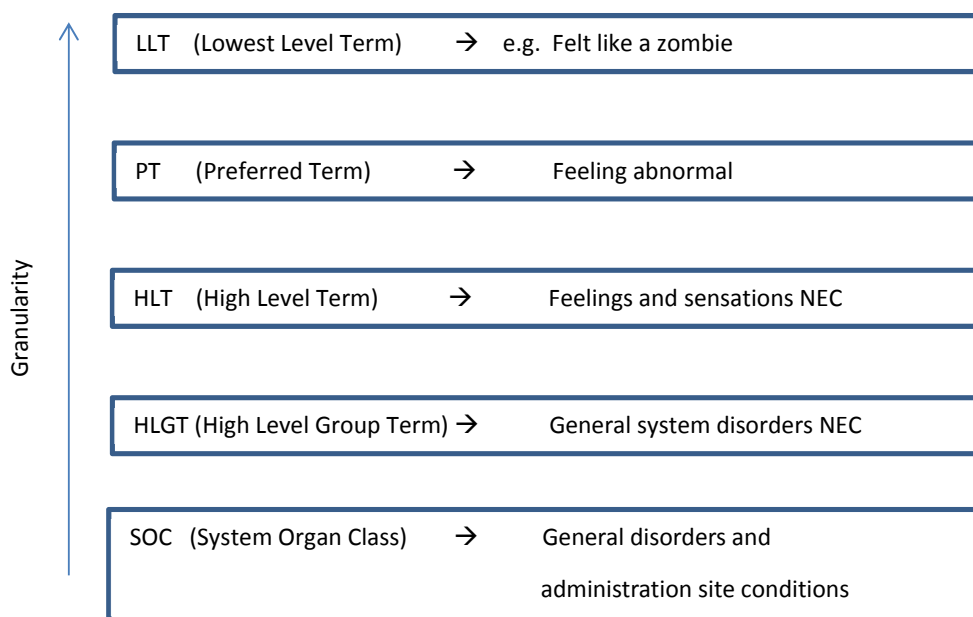
MedDRA

After a short introduction to MedDRA the handling of a recoding or better refreshing process will be presented and how to use the result to update the existing AE dataset and to create in a final step a so called adverse event history domain that contains the former version of the MedDRA related variables.

The MedDRA terminology was developed as a medically validated medical terminology for utilization throughout the regulatory process. The developers of the terminology designed a structure that promotes specific and comprehensive data entry and flexible data retrieval. Each term in MedDRA has a unique non-expressive code, "meaning no sense".

Adverse Events are collected in textual or verbatim content, which is processed through a coding dictionary so that similar verbatim content is grouped together by classifying them into a hierarchy of medical granularity, as the following figure illustrates:

The structural hierarchy of the MedDRA Terminology is simplified represented as:



PhUSE 2014

MedDRA Update

Based on Standard Operation Procedures, operation manuals and in cooperation with a centralized global Coding team - that provides and maintains standards and tools - an updating to the controlled vocabulary of the existing thesauri versions is initialized. To achieve homogenous, consistent and medical accurate coding output a macro is used for this so-called REFRESH process for in house coded studies in an integrated database, that is based on the clinical studies analysis data sets. The up-versioning is done by unique terms for the selected thesaurus as the combination of the existing verbatim (AETERM), the modified verbatim (AEMODIFY) and the clarified verbatim (AETERMCV). No duplicated information is used and no additional subject information is required. A consistency is reached as individual terms are coded initially identically in a single clinical study and also across several studies in an integrated data base.

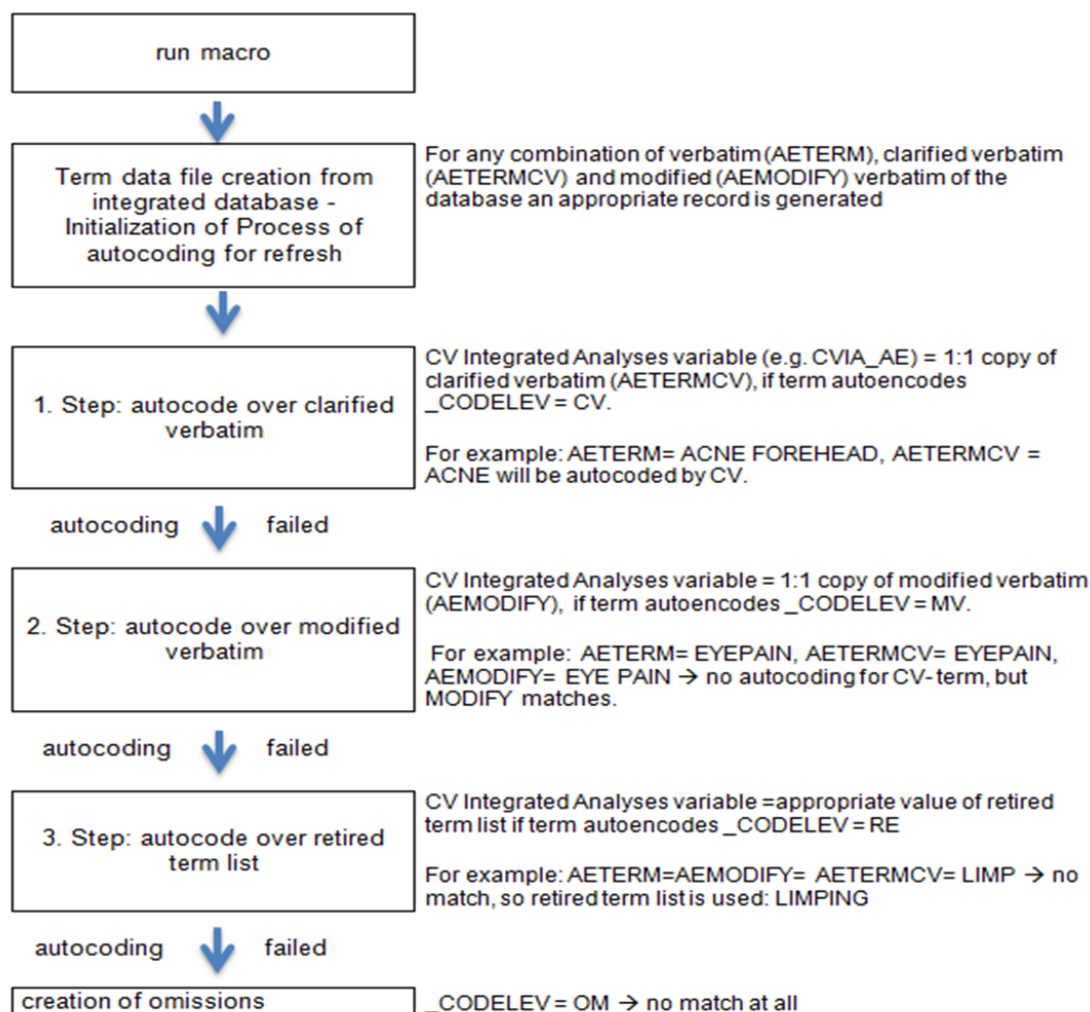
An initial SAS® macro provided by programming gives the necessary information to the coding environment so that their internal macros can update the requested dictionary. The now described macro parameters are requested to ensure that an update is successfully initialized and that different input datasets and thesauri can be addressed.

SOURCE_DIR	= data directory
SOURCE_DATA_SET	= ADAE
TERM_DATA_SET	= ADAExxx
REPORT_DIRECTORY	= RESULTS
IA_CREATE_OMISSIONS	= Y/N
IA_CREATE_TRR	= Y/N → a Term Review Report can be created as datasets and as Microsoft Office® Word and Excel files for further documentation and investigation
ORIGINAL_TERM_VARIABLE	= AETERM
CLARIFIED_TERM_VARIABLE	= AETERMCV
MODIFIED_TERM_VARIABLE	= AEMODIFY
WORK_FLOW_VARIABLE	= AETERMWF
CODE_VARIABLE_1	= AEMLLT
CODE_VARIABLE_2	= needed for Concomitant Medication WHO DD coding
CODE_VARIABLE_3	= needed for Concomitant Medication WHO DD coding
CODING_THESAURUS	= meddra
IA_IDENT	= &project.
TITLE1_REPORT	= .ADVERSE EVENT TERMS .

In addition the coding macros have to be included.

In a several step approach the existing verbatims will be checked against the current thesaurus. In a first step it is looked for all terms that autocode. This means no further action is required and those terms are updated including the new lowest level term code.

PhUSE 2014



Unfortunately there are terms that cannot be autocoded. These failures are named omissions and are delivered to the coding environment, from where they have to be coded manually by medical coders after medical review. When this task is successfully completed the macro has to rerun to achieve a complete updated dataset that contains the necessary information for an update to the thesaurus version for the existing dataset.

After the coding process is finalized a so called Term data file is available and the encoding information can be used for further processing and has the following structure:

Variables	Description	
AETERM	Reported Term for the Adverse Event	Original verbatim in database
AETERMCV	Clarified Verbatim	Original clarified verbatim in database
AEMODIFY	Modified Reported Term	Original modified verbatim in database
AETERMWF	WORKFLOW STEP FOR ADVERSE EVENT	Workflow variable
CVIA_AE	CLARIFIED VERB. FOR IA	New clarified verbatim for recent version and future updates

PhUSE 2014

MODIA_AE	MODIFIED VERB. FOR IA	New modified verbatim for recent version and future updates
AEMLLT	MedDRA Term Code	New AEMLLT in most recent MedDRA version
_CODELEV	FLAG for CODING LEVEL(CV, MV, RE or OM)	Which level is used for update
QUERY_XXXX		Variables for coding function related to the coding process
THES_VERSION	THESAURUS VERSION	Most recent version, e.g. 17.0
UPDATE_D	DAY TERM WAS ADDED TO TERM DATA SET	
UPDATE_M	MONTH TERM WAS ADDED TO TERM DATA SET	
UPDATE_Y	YEAR TERM WAS ADDED TO TERM DATA SET	

This received Term data file can be further processed to update the ADAE data set to the most recent version and to store the old ADAE dataset in a history domain.

WORKFLOW TO UPDATE DATASETS

Seeing the difficulties and challenges as described before, Bayer Integrated Analyses has chosen an approach that creates two datasets:

- 1.) An updated ADAE dataset containing the most recent MedDRA version
- 2.) A so-called AE history dataset ADAEHIST storing the prior dictionary version(s)

Therefore a macro can be used that uses the information from the coding process and enriches the ADAE dataset with the complete MedDRA hierarchy. In this step it is not intended to add MedDRA related variables like SMQ/CMQ's, because they can vary over time and events of interest might be defined as combinations of e.g. SMQ's in a temporary context. But nevertheless a further iteration can be added that accomplishes this task.

It should be mentioned as the structure of the medical history dataset is quite similar to the adverse event dataset and as a consequence this workflow is also suitable for this domain. For other thesauri like the WHO Drug Dictionary for groupings of concomitant medications the same approach can be used. The names of the needed variables are slightly different, but the work flow is the same.

Following basic parameters are needed:

DATASET : dataset to be updated
 PREFIX_ : prefix for MedDRA-Variables
 TERMDATA : dataset containing new codes (including library name)
 OLDVERSION : old MedDRA version
 REPORTTERM : variable containing reported term AETERM

PhUSE 2014

CLARIFIEDTERM : variable containing clarified term AETERMCV

MODIFIEDTERM : variable containing modified term AEMODIFY

CODEVARIABLE : variable containing updated code AELLT

APPEND : if history file already exists =y to append to history file

Macro variables for paths/ library names of datasets might be added. But as the paths for the required datasets normally not change it is not intended to add them to basic parameters.

The macro initially creates or appends an existing history dataset based on the existing ADAE. This can easily be done when initially setting the current ADAE and rename it to ADAEHIST. When this dataset already exists the ADAE will be set together with the available ADAEHIST. The only additional step is that the prior MedDRA version must be added as additional variable to the dataset, so that a distinction between different dictionary versions is possible.

In the next step the macro updates the original dataset when using PROC SQL and merging via a left join the data term file containing the new information to the ADAE dataset. The base information for AETERM, AEMODIFY and AETERMCV are the same and so the merge is done using these three variables. A requirement is to rename the new LLT- code variable from the update information. Then the AEMODIFY variable is updated using the result from the coding refresh. After this is done, the complete (primary) hierarchy (i.e. PT, HLT, HLG, SOC) is added to the dataset via attaching the corresponding MedDRA formats to the (new) LLT code variable.

CONCLUSION

The advantages of this workflow are that it in terms of usability of the analysis dataset. Especially ADAE stays well-arranged and does not exceed the allowed size for electronic submissions to the FDA. In therapeutic areas like Oncology this can quite easily happen. For example in one of our last submissions 47 SMQs and CMQs were needed for adverse event reporting. When you add all the code variables, numeric versions etc. you double at least this number and this with every new MedDRA version. For example: one of our oncology databases started with MedDRA version 4.0 and when you would have done this from the beginning then there would be with 25 MedDRA versions so far multiplied with 94 variables an additional 2350 variables in the database.

Optionally – depending on the requirements for an ongoing submission – the former MedDRA versions can be added to the ADAE dataset, but through a life-cycle management this is in the author's opinion not recommendable.

Analyses based on former versions can be reproduced without changing of existing programs. Only a macro parameter for the ADAE(HIST) dataset and the required MedDRA version is needed.

Furthermore the requirement of traceability can be achieved, when comparing the datasets respectively the different dictionary versions. This can be supported by a difference file or listing.

An improvement that can be achieved in the current workflow for future releases is that the complete MedDRA hierarchy is updated and can be added in the term data set. When in addition to the most recent lowest level term also the preferred term, the high level term, the high level group term and finally the system organ class with their respective codes and decodes are available after the up-versioning process they can be easily attached to the ADAE dataset.

PhUSE 2014

REFERENCES

^aAnalysis Data Model (ADaM) Data Structure for Adverse Event Analysis Version 1.0. Available from:

http://www.cdisc.org/system/files/all/standard_category/application/pdf/adam_ae_final_v1.pdf

[Accessed 29 August 2014]

^bICH HARMONISED TRIPARTITE GUIDELINE – Clinical Safety Data Management: Definitions and Standards for Expedited Reporting – E2A. Available from:

http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf

[Accessed 29 August 2014]

^cMedical Dictionary for Regulatory Activities (MedDRA). Available from:

<http://www.meddra.org/>

http://www.meddra.org/sites/default/files/guidance/file/intguide_17_0_english.pdf

http://www.meddra.org/sites/default/files/guidance/file/smq_intguide_17_0_english.pdf

[Accessed 29 August 2014]

^dAnalysis Data Model (ADaM) Implementation Guide Version 1.0. Available from:

http://www.cdisc.org/system/files/all/standard/application/pdf/adam_implementation_guide_v1.0.pdf

[Accessed 29 August 2014]

^eAnalysis Data Model Structure for Occurrence Data Version 1.0 Draft. Available from:

<http://portal.cdisc.org/CT/Review%2520Documents/ADaM->

[ODS_v1.0draft.pdf&rct=j&frm=1&q=&esrc=s&sa=U&ei=OjYEVm_iGI707Aba94DoBQ&ved=0CBcQFjAA&usg=AFQjCNE5oAsD9FBWapRe8FcRvoaS3EMG5w](http://portal.cdisc.org/CT/Review%2520Documents/ADaM-ODS_v1.0draft.pdf&rct=j&frm=1&q=&esrc=s&sa=U&ei=OjYEVm_iGI707Aba94DoBQ&ved=0CBcQFjAA&usg=AFQjCNE5oAsD9FBWapRe8FcRvoaS3EMG5w)

[Accessed 29 August 2014]

^fUS Food and Drug Administration: Study Data Technical Conformance Guide. Available from:

<http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>

[Accessed 29 August 2014]

ACKNOWLEDGMENTS

I would like to thank Sabine Fiala-Buskies, Jörg Güttner, Jean-Marleen Majewski and Marit Weiss (all Bayer Pharma AG) for their valuable input, time and contributions towards this paper.

Brand and product names are trademarks of their respective companies.

CONTACT INFORMATION

Comments and questions are valued. Contact the author at:

Peter Bonata

Bayer Pharma AG

Aprather Weg 18a

D-42113 Wuppertal

Germany

Phone: +49 202 365263

Email: peter.bonata@bayer.com

Web: <http://www.bayer.com>